

Analisis Sentimen Terhadap Opini Publik pada Sosial Media Twitter Menggunakan Metode Support Vector Machine

Ade Dwi Dayani[✉], Yuhandri, Gunadi Widi Nurcahyo

Fakultas Ilmu Komputer, Universitas Putra Indonesia YPTK, Jl. Raya Lubuk Begalung, Padang, 25221, Indonesia

[✉] Adedwidayani@gmail.com

Abstract

The development of the topic of childfree has now become a hot topic of discussion among the public. The reemergence of the topic of childfree on social media is due to a social media activist who decided to choose childfree and put it forward on social media. This research aims to carry out sentiment classification analysis of public opinion on Twitter social media. The classification analysis method used adopts the performance of the Support Vector Machine (SVM) method to provide optimal output. The research dataset was taken using crawling techniques sourced from social media Twitter. The research dataset obtained will be classified into positive, negative, and neutral sentiment models. The results of SVM analysis testing based on sample data obtained classification analysis results with an accuracy level of 69.69%, recall of 45.60%, precision of 51.56%, and F1-Score of 46%. Based on these results, the performance of SVM analysis shows quite good performance in carrying out classification analysis of public opinion on Twitter social media. This research can contribute to providing new knowledge in classification using the Support Vector Machine method and seeing how the childfree topic develops on Twitter social media in Indonesia.

Keywords: Childfree, Sentiment Analysis, Public Opinion, Social Media, Support Vector Machine

Abstrak

Perkembangan topik *childfree* kini menjadi pembahasan yang ramai diperbincangkan oleh publik. Hangatnya kembali topik *childfree* di karenakan seorang pegiat sosial media yang memutuskan untuk memilih *childfree* dan mengemukakannya ke media sosial. Penelitian ini bertujuan untuk melakukan sentimen analisis klasifikasi terhadap opini publik pada sosial media twitter. Metode analisis klasifikasi yang digunakan mengadopsi kinerja metode Support Vector Machine (SVM) untuk menyajikan keluaran yang optimal. Dataset penelitian diambil dengan menggunakan teknik *crawling* yang bersumber dari sosial media Twitter. Dataset penelitian yang diperoleh akan di klasifikasi ke dalam model sentimen positif, negatif, dan netral. Hasil pengujian analisis SVM berdasarkan data sampel diperoleh hasil analisis klasifikasi dengan tingkat akurasi sebesar 69,69%, *recall* sebesar 45,60%, *precision* sebesar 51,56%, dan *F1-Score* 46%. Berdasarkan hasil penelitian yang diperoleh, kinerja analisis SVM menunjukkan performa yang cukup dalam melakukan analisis klasifikasi terhadap opini publik pada sosial media twitter. Penelitian ini dapat berkontribusi dalam memberikan pengetahuan baru dalam pengklasifikasian menggunakan metode Support Vector Machine serta melihat bagaimana perkembangan topik *childfree* pada media sosial Twitter di Indonesia.

Kata kunci: Childfree, Analisis Sentimen, Opini Publik, Media Sosial, Support Vector Machine

KomtekInfo is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.



1. Pendahuluan

Media sosial adalah sebuah platform di mana pengguna dapat berinteraksi secara online tanpa adanya batasan ruang dan waktu, memungkinkan mereka untuk bersosialisasi dan berinteraksi satu sama lain. Salah satu pemanfaatan media sosial yaitu untuk mengungkapkan dan berbagi pendapat, pandangan, usul, bahkan kritik terhadap suatu topik, isu, berita dan produk yang dibagikan ke ruang public [1]. Masyarakat bebas memberikan tanggapan dan pendapatnya di berbagai media sosial dan menjadikan media sosial tempat untuk mencurahkan segala hal [1]. Twitter menjadi salah satu platform media sosial yang populer digunakan untuk menyampaikan pendapat [2].

Twitter merupakan situs *microblogging* yang dilambangkan dengan burung berwarna biru dan biasanya digunakan untuk membagikan apapun itu melalui cuitan atau tweet [3]. Salah satu fitur yang ditawarkan oleh Twitter yaitu, fitur tagar (*hashtag*) dan *trending topic* [4]. Fitur ini membantu pengguna untuk mengikuti topik, berita, fenomena, peristiwa dan percakapan yang mereka minati ataupun yang sedang ramai diperbincangkan [4]. Salah topik yang sedang ramai diperbincangkan adalah *childfree* [5].

Childfree adalah sebuah keputusan yang diambil oleh seseorang atau pasangan untuk tidak memiliki keturunan dalam pernikahan mereka [6]. Keputusan ini dipilih dengan penuh kesadaran, baik laki-laki maupun perempuan, memiliki hak untuk memilihnya untuk

menentukan pilihan hidup mereka [7]. Banyak hal yang dijadikan alasan oleh seseorang untuk memilih Childfree di dalam kehidupannya, seperti kekhawatiran dalam perkembangan sang anak, masalah pribadi, ekonomi, bahkan faktor lingkungan sekitarnya [7].

Analisis sentimen dilakukan untuk melihat cara pandang atau bagaimana seseorang berpendapat terkait suatu masalah atau objek, apakah cenderung berpandangan atau beropini negatif atau positif [8]. Positif berarti adanya aspek suka, setuju, dan sejenisnya terhadap suatu topik. Sebaliknya, negatif memiliki arti ketidaksukaan, ketidaksetujuan, kekecewaan, dan lainnya, kemudian netral berarti berada diantaranya [7]. Analisis sentimen menjadi salah satu teknik yang banyak digunakan untuk memahami pendapat dan perasaan seseorang berdasarkan pada data teks [9].

Natural Language Processing (NLP) merupakan salah satu bidang ilmu di dalam bidang *artificial intelligent* (AI) [10]. Dalam menganalisis sentimen, NLP diperlukan untuk meningkatkan kualitas bahasa dalam komentar ulasan agar mencapai hasil analisis yang optimal [11]. NLP juga digunakan untuk klasifikasi teks dengan menggunakan metode *machine learning*, salah satunya Support Vector Machine [12].

Support Vector Machine (SVM), dikenal sebagai teknik *machine learning* paling mutakhir dibandingkan dengan teknik *machine learning* lainnya [13]. SVM mampu mengidentifikasi dan berhasil digunakan dalam pengenalan pola [13]. Metode SVM sering digunakan dalam analisis sentimen sebagai alat untuk mengkategorisasikan teks ke dalam kategori positif, negatif dan netral [14].

Salah satu analisis sentimen dalam menganalisa opini masyarakat terkait vaksin Covid-19 di Twitter menggunakan metode SVM dan Naïve Bayes, menyajikan hasil analisis dengan tingkat *accuracy*, *precision*, dan *recall* sebesar 90,47%, 90,23%, dan 90,78% [1]. Perbandingan empat metode pada isu kebocoran data kartu identitas ponsel menggunakan metode Logistic Regression, Support Vector Machine (SVM), Random Forest, dan model IndoBERT menyajikan hasil keluaran SVM dengan performa terbaik dilihat dari nilai *f1-score* sebesar 0.81 [15].

Penelitian lainnya juga menjelaskan bahwa performa metode SVM memberikan hasil terbaik dalam melakukan analisis sentimen dengan akurasi sebesar 77.4% [9]. Analisis sentimen juga dilakukan untuk melihat antusiasme publik tentang dirilisnya produk Apple dipasaran menggunakan metode SVM dengan menunjukkan hasil yang memuaskan berdasarkan nilai akurasi, presisi dan *f1-score*, yaitu 89,21%, 92,43%, dan 93,95% [16]. Analisa sentimen juga dilakukan untuk melihat respon publik terhadap kebijakan yang menggunakan beberapa metode seperti Decision Tree,

Naïve Bayes, Random Forest, dan Support Vector Machine (SVM) menunjukkan bahwa SVM mengungguli metode lainnya dengan nilai akurasi, presisi dan *f1-score* sebesar, yaitu 0.99, 0.92, dan 0.95 [17]. Analisis sentimen dengan metode SVM juga dapat dilakukan untuk melihat bagaimana tanggapan pengguna aplikasi Gojek terhadap pelayanan yang di berikan oleh Gojek di media sosial Twitter memberikan hasil dengan tingkat akurasi sebesar 79,19% dan akurasi kappa sebesar 21% [8]. Penelitian lainnya juga menjelaskan bahwa metode SVM digunakan untuk analisis sentimen publik bjorka memberikan hasil akurasi sebesar 62,33% serta lebih cenderung ke netral [18].

Berdasarkan penjelasan beberapa penelitian sebelumnya maka proses analisis sentimen klasifikasi juga dapat dilakukan terhadap opini publik pada sosial media twitter. Proses analisis tersebut dilakukan dengan menggunakan metode SVM pada konsep Natural Language Processing (NLP) untuk menentukan model sentimen positif, negatif, dan netral terhadap dataset penelitian. Berdasarkan hasil performa metode SVM dalam proses analisis maka penelitian ini bertujuan untuk melihat hal apa yang sedang terjadi di tengah masyarakat terkait *childfree* kemudian bagaimana kecenderungan opini publik terhadap isu *childfree* yang dituangkan pada media sosial Twitter. Kontribusi yang disajikan dalam penelitian ini dapat memberikan pengetahuan baru dalam pengklasifikasian menggunakan metode Support Vector Machine serta melihat bagaimana perkembangan topik *childfree* pada media sosial Twitter di Indonesia.

2. Metodologi Penelitian

Metodologi penelitian adalah tahapan yang menjadi acuan yang diterapkan proses analisis sentimen. Proses analisis sentimen dilakukan dengan melewati dua fase, yaitu *Natural Language Processing* (NLP) dan pengklasifikasian dengan metode *machine learning*. *Natural Language Processing* (NLP) melibatkan penggunaan algoritma *machine learning* untuk mempelajari pola dari data bahasa alami yang telah diberikan [12]. Algoritma ini dapat digunakan untuk tugas-tugas seperti klasifikasi teks, ekstraksi informasi, pemodelan, dan terjemahan mesin [12]. Algoritma yang digunakan pada penelitian ini adalah Support Vector Machine (SVM).

2.1 NLP (Natural Language Processing)

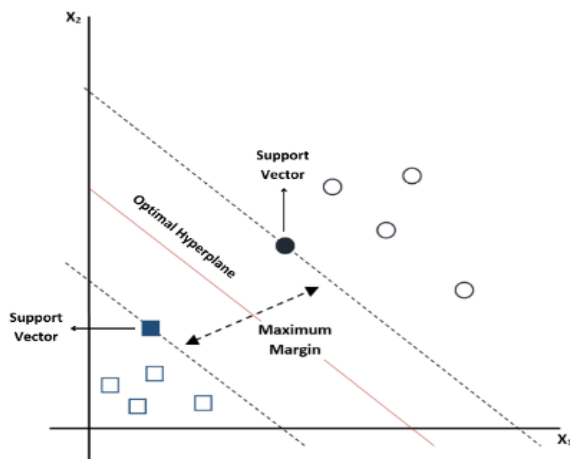
NLP berfokus pada pengolahan bahasa alami, bahasa yang umum digunakan oleh manusia untuk berkomunikasi satu sama lain, sehingga memungkinkan mesin untuk apa yang diinginkan oleh pengguna [10]. NLP diperlukan untuk memperbaiki bahasa pada isi komentar atau tweet sehingga mencapai kinerja maksimal dalam analisis sentimen [11]. NLP dalam konsep *machine learning* dapat menghasilkan data yang

diberi nilai sentimen, dilakukan proses *preprocessing* data, yaitu :

- a. *Case Folding*, tahapan di mana mengubah keseluruhan bentuk huruf pada sebuah teks dokumen ke dalam huruf kecil.
- b. *Cleaning*, berfungsi untuk menghapus atau menghilangkan data dari suatu kalimat yang memiliki unsur simbol atau kata yang tidak dibutuhkan, misalnya *username* (@username), *hashtag* (#), emoji, URL, *email* dan kata kunci.
- c. *Tokenizing*, berfungsi untuk memotong atau memisahkan suatu kalimat ke dalam satu kata. *Tokenizing* memotong kalimat menjadi kata, karakter, simbol, atau tanda baca, sehingga menjadi kata yang selanjutnya dapat dianalisis.
- d. *Stopword Removal*, dilakukan dengan tujuan untuk menghapus kata yang tidak bermakna sesuai dengan kata yang terdapat dalam daftar *stopword*. *Stopword* adalah kumpulan kata yang tidak penting dan tidak akan digunakan.
- e. *Stemming*, bertujuan untuk menghapus imbuhan-imbuhan dalam kata, baik di awal maupun akhir kata dengan memperhatikan awalan umum dan sufiks (imbuhan yang melekat pada kata dasar), yang dapat ditemukan dalam kata.

2.2 Support Vector Machine

SVM memecahkan masalah klasifikasi dengan mencari fungsi pemisah (*hyperplane*) margin maksimum yang dapat memisahkan dua kelas secara optimal [19]. *Hyperplane* dengan margin yang lebih besar lebih akurat dalam mengklasifikasikan data dibandingkan yang lebih kecil, hal ini dikenal dengan istilah Maximum Marginal *Hyperplane* [20]. Konsep *hyperplane* disajikan pada Gambar 2.



Gambar 2. Konsep Hyperplane pada SVM

Margin adalah dua kali jarak antara *hyperplane* dan *support vector*, dimana *support vector* adalah titik yang berada paling dekat dengan *hyperplane* [19]. *Support vector* juga dapat dikatakan objek-objek data terluar yang paling dekat dengan *hyperplane*. *Support vector*

inilah yang nantinya akan diperhitungkan oleh SVM [8].

Pada dasarnya, SVM merupakan suatu *linear classifier*, namun SVM dapat dikembangkan menjadi *nonlinear classifier*. Konsep kernel trick pada ruang berdimensi lebih tinggi digunakan untuk menangani *nonlinear classifier* [21]. Beberapa fungsi kernel *trick* pada SVM yang sering digunakan, yaitu:

- a. Kernel Linear

Kernel *linear* sering digunakan untuk data dua kelas atau data yang diklasifikasikan secara *linear* pada Persamaan 1.

$$K(x_i, x_j) = x_i \cdot x_j \quad (1)$$

- b. Kernel Radial Bias Function (RBF)

Kernel RBF atau *Gaussian* digunakan untuk menyelesaikan masalah data yang tidak terpisah secara *linear* atau data dengan banyak kelas Persamaan 2.

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (2)$$

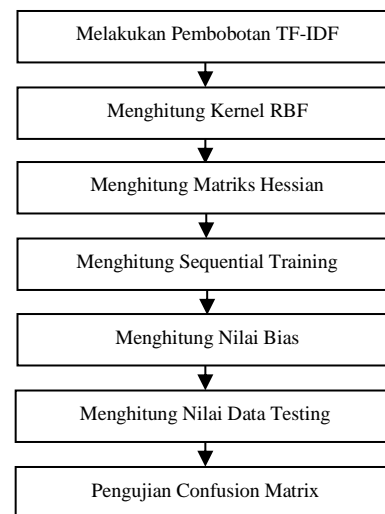
- c. Kernel Polynomial

Kernel *polynomial* merupakan kernel *trick* lainnya yang digunakan untuk menyelesaikan masalah data yang tidak terpisah secara *linear* pada Persamaan 3.

$$K_{x_i x_j} = (x_i \cdot x_j + c)^d \quad (3)$$

K merupakan fungsi kernel. x_i, x_j merupakan dokumen dari *dataset*. γ merupakan parameter kecepatan proses *training* atau nilai gamma. \exp adalah basis dari logaritma alami, dan d adalah derajat *polynomial*.

Proses klasifikasi atau perhitungan menggunakan metode Support Vector Machine, dilakukan dengan pengujian data *training* dan data *testing*. Data *training* dihitung menggunakan metode *Sequential Minimal Optimization* (SMO) [21]. Adapun tahapan pada metode SVM dapat dilihat pada Gambar 3.



Gambar 3. Tahapan Metode SVM

Proses implementasi metode Support Vector Machine akan dimulai dengan menggunakan hasil dari TF-IDF yang telah didapatkan pada proses sebelumnya. Metode Support Vector Machine memiliki tahapan – tahapan dalam proses perhitungannya, yaitu:

- a. Melakukan pembobotan TF-IDF
Sebelum mengimplementasikan metode SVM dan setelah *preprocessing* data, dilakukan terlebih dahulu pembobotan kata menggunakan TF - IDF. Hasil dari TF - IDF kemudian digunakan untuk pengimplemen-tasian metode SVM. TF-IDF merupakan proses pembobotan dari setiap kata (term) yang terdapat dalam dokumen [22]. TF-IDF akan melakukan perhitungan *tf*, *df*, *idf*, dan TF-IDF. Menghitung TF-IDF dilakukan dengan Persamaan 4 & 5.

$$W_{dt} = TF_{dt} \times IDF_{ft} \quad (4)$$

$$idf_t = \log\left(\frac{N}{df_t}\right) \quad (5)$$

W_{dt} merupakan bobot TF-IDF atau nilai TF-IDF yang akan dicari. TF_{dt} merupakan jumlah frekuensi kata yang muncul dalam sebuah dokumen. IDF_{ft} merupakan jumlah inverse frekuensi dokumen tiap kata. Df_t merupakan jumlah frekuensi dokumen tiap kata dan N merupakan jumlah total dokumen.

- b. Melakukan perhitungan kernel
Kernel yang digunakan pada penelitian ini adalah kernel RBF.
- c. Menghitung matrix hessian dengan Persamaan 6.

$$D_{i,j} = y_i, y_j (K(x_i, x_j) + \lambda^2) \quad (6)$$

$D_{i,j}$ merupakan elemen matriks ke ij (data ke-1 sampai data ke-n). Y_i merupakan kelas atau label sentimen data i. Y_j merupakan kelas atau label sentimen data j. Kemudian, λ^2 merupakan batas teoritis yang diturunkan

- d. Menghitung Sequential Training
Sequential Training merupakan tahapan di mana memproses data *training* pada SVM. Proses ini memiliki beberapa tahapan yang dilakukan seperti melakukan inialisasi beberapa variabel, yaitu α_i , γ , C , dan ε serta menghitung E_i , $\delta\alpha_i$, dan α . Perhitungan nilai *error rate* disajikan dalam Persamaan 7.

$$E_i = \sum_j^i a_j D_{ij} \quad (7)$$

E_i merupakan nilai *error* data ke i. Tujuan dari perhitungan nilai *error rate* adalah untuk mendapatkan hasil *alpha* terbaru dari setiap data *training*. Setelah nilai *error rate* didapatkan, maka

nilai *error rate* akan digunakan untuk menghitung nilai delta alfa ($\delta\alpha_i$). Perhitungan nilai *delta alpha* disajikan dalam persamaan berikut:

$$\delta\alpha_i = \min(\max[\gamma(1 - E_i), \alpha_i], C - \alpha_i) \quad (8)$$

$\delta\alpha_i$ merupakan delta α ke i. γ merupakan nilai gamma. Kemudian C merupakan *complexity*. Nilai dari delta alfa kemudian digunakan untuk mencari nilai alfa baru yang akan digunakan untuk menghitung nilai *error rate* selanjutnya. Perhitungan nilai *alpha* baru disajikan dalam persamaan berikut:

$$\alpha_i \text{ baru} = \alpha_i + \delta\alpha \quad (9)$$

- e. Menghitung Nilai Bias
Perhitungan nilai bias disajikan dalam Persamaan 10-12 sebagai berikut:

$$w \cdot x^+ = \alpha_i Y_i K(w \cdot x^+) \quad (10)$$

$$w \cdot x^- = \alpha_i Y_i K(w \cdot x^-) \quad (11)$$

$$b = \frac{1}{2}(w \cdot x^+ + w \cdot x^-) \quad (12)$$

$w \cdot x^+$ merupakan nilai kernel data x dengan data x kelas positif yang memiliki nilai α tertinggi. $w \cdot x^-$ merupakan nilai kernel data x dengan data x kelas negatif yang memiliki nilai α tertinggi.

- f. Menghitung Nilai Data Testing
Perhitungan nilai data *testing* disajikan dalam Persamaan 13.

$$f(x) = \text{sign} \sum_{i=0}^n (\alpha_i y_i K(x, x_i) + b) \quad (13)$$

x merupakan titik data masukan Support Vector Machine. α_i merupakan nilai bobot setiap titik data. b adalah parameter hyperplane yang dicari (nilai bias) dan $K(x, x_i)$ adalah fungsi kernel data testing.

- g. Pengujian Confusion Matrix
Confusion Matrix digunakan untuk mengukur tingkat akurasi dan kinerja dari metode SVM yang diterapkan. Metode evaluasi yang digunakan mencakup akurasi, presisi, *recall*, dan *f1-score*. Rumus dari metode evaluasi disajikan dalam Persamaan 14-17.

$$\text{accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (14)$$

$$\text{precision} = \frac{TP}{TP+FP} \quad (15)$$

$$\text{recall} = \frac{TP}{TP+FN} \quad (16)$$

$$f1 \text{ score} = \frac{2 \times (\text{recall} \times \text{precision})}{(\text{recall} + \text{precision})} \quad (17)$$

2.3 Dataset Penelitian

Dataset penelitian dikumpulkan dengan teknik *crawling* dari sosial media twitter berdasarkan rentang waktu dari 1 Februari 2023 sampai dengan 1 April 2023. Dataset tersebut diambil berdasarkan kata kunci dari *tweet* yang mengandung "Childfree". Proses pengambilan dataset pada aplikasi sosial media twitter menggunakan *python* dan memperoleh data sebanyak 500 *tweet*. Adapun sampel dataset penelitian dapat disajikan pada Tabel 1.

Tabel 1. Data Twitter Opini Childfree

Username	Tweet
sanctuari	Btw, setahuku feminis itu bukan ngajarin untuk jadi benci pria, childfree, ga menikah, dll. Feminis memperjuangkan hak wanita untuk bebas menentukan jalan hidupnya tanpa didikte siapapun.
urpretiers	Kok bisa ya topik obrolan gue sama temen gue tuh sampe ke 'sapi childfree' $\delta\ddot{Y}\ll\Upsilon\delta\ddot{Y}\sim/f\delta\ddot{Y}^{\text{TM}}$,
...	...
yuliushr	@bininyakjongin @kacanggodhog @tanyakanrl Gausah ngajak orang lain buat mati sendirian ditemenin sama kucing, sesat nih orang childfree"
CNNIndonesia	@kegblgnunfaedh Makanya gw setuju childfree diterapin disini karena mayoritas ortu ortunya pada gblk gabisa didik anaknya

Tabel 1 merupakan data yang diperoleh dari proses *crawling* menggunakan Python. data yang telah didapat kemudian di *export* ke Excel dengan format *.csv* untuk selanjutnya dilakukan *preprocessing* data. Data yang didapat masih berupa data mentah yang tidak beraturan dan tidak terstruktur.

3. Hasil dan Pembahasan

3.1 Preprocessing Data

Dataset yang telah dikumpulkan dari proses *crawling* adalah data yang masih berbentuk kata atau kalimat tidak terstruktur dan tidak beraturan, maka membutuhkan proses yang disebut *preprocessing* agar menghasilkan data yang bersih sehingga hasil analisis sentimen menjadi lebih akurat. *Preprocessing* dilakukan dengan menggunakan *python*. Tahapan *preprocessing* dalam penelitian ini adalah:

- a. *Case Folding*, tahapan di mana mengubah keseluruhan bentuk huruf pada sebuah *tweet* menjadi huruf kecil. Hasil dari *case folding* disajikan pada Tabel 2.

Tabel 2. Tabel Proses Case Folding

Sebelum	Sesudah
@recehtapisayng Anaknya nanti kalau ngerantau gak bisa ngomong "kangen masakan mama" Eh tapi siapa tau mereka childfree.	@recehtapisayng anaknya nanti kalau ngerantau gak bisa ngomong "kangen masakan mama" eh tapi siapa tau mereka childfree.

- b. *Cleaning*, berfungsi untuk menghapus atau menghilangkan menghapus tanda baca seperti [!'"#\$%&'()*+,-./:;=>?@[\\]^_`{|}~] yang dapat dilihat pada Tabel 3.

Tabel 3. Proses Cleaning

Sebelum	Sesudah
@recehtapisayng anaknya nanti kalau ngerantau gak bisa ngomong "kangen masakan mama eh tapi	anaknya nanti kalau ngerantau gak bisa ngomong kangen masakan mama eh tapi
masakan mama" eh tapi siapa tau mereka childfree.	siapa tau mereka childfree

Tokenizing, berfungsi memotong kalimat menjadi kata, karakter, simbol, atau tanda baca, sehingga menjadi kata yang selanjutnya dapat dianalisis. Proses ini dilakukan dengan menggunakan *library* fungsi *RegexTokenizer()* dari *library nltk*.

Tabel 4. Proses Tokenizing

Sebelum	Sesudah
anaknya nanti kalau ngerantau gak bisa ngomong kangen masakan mama eh tapi siapa tau mereka childfree	['anaknya', 'nanti', 'kalau', 'ngerantau', 'gak', 'bisa', 'ngomong', 'kangen', 'masakan', 'mama', 'eh', 'tapi', 'siapa', 'tau', 'mereka', 'childfree']

Stopword Removal, dilakukan untuk menghilangkan kata yang tidak penting atau tidak memiliki arti. Proses ini menggunakan fungsi *library* Sastrawi, yaitu *stopWordRemoverFactory*.

Tabel 5. Stopword Removal

Sebelum	Sesudah
['anaknya', 'nanti', 'kalau', 'ngerantau', 'gak', 'bisa', 'ngomong', 'kangen', 'ngomong', 'kangen', 'masakan', 'mama', 'eh', 'masakan', 'mama', 'eh', 'tapi', 'siapa', 'tau', 'mereka', 'childfree']	['anaknya', 'ngerantau', 'ngomong', 'kangen', 'masakan', 'mama', 'eh', 'tau', 'childfree']

Stemming, bertujuan untuk menghilangkan kata imbuhan. Proses ini menggunakan *library* Sastrawi. Sastrawi digunakan untuk mengubah kata berimbuhan bahasa Indonesia menjadi bentuk dasarnya.

Tabel 6. Proses Stemming

Sebelum	Sesudah
['anaknya', 'ngerantau', 'ngomong', 'kangen', 'masakan', 'mama', 'eh', 'tau', 'childfree']	['anak', 'ngerantau', 'ngomong', 'kangen', 'masakan', 'mama', 'eh', 'tau', 'childfree']

3.2 Pembagian dan Pelabelan Data

Total *dataset* yang didapat setelah melakukan *preprocessing* dan menghilangkan *tweet* yang memiliki duplikat adalah sebanyak 461 data. Langkah selanjutnya adalah pembagian data, di mana data dibagi menjadi dua, yaitu Data Training dan Data Testing, dengan rasio 80 : 20.

- a. Data *training* digunakan untuk pembuatan model *machine learning*, dengan menggunakan label positif, negatif, dan netral.

b. Data *testing* digunakan untuk pengujian dan implementasi metode Support Vector Machine. Pelabelan pada data *tweet* untuk data *training* dilakukan secara manual dengan seorang pakar. Tweet dibaca satu per satu untuk mengetahui apakah *tweet* tersebut bernada positif, negatif dan netral.

3.3 TF-IDF

a. Perhitungan Nilai tf dan df

Nilai *tf*, didapat dengan melihat frekuensi atau banyaknya sebuah atau kata muncul di dalam sebuah dokumen. Apabila sebuah kata muncul maka dilambangkan dengan angka 1, jika kata tersebut tidak ada pada dokumen maka dilambangkan dengan angka 0. Nilai *df*, didapat dengan menghitung jumlah dokumen di mana *term* tersebut muncul dengan minimal satu dokumen pada Tabel 7.

Tabel 7. Perhitungan Nilai *tf* dan *df*

Term	TF									DF
	Data Training					Data Testing			DF	
	D	D	D	D	D	U	U	U		
1	2	3	4	5	1	2	3			
tahu	1	0	0	0	0	0	0	0	1	
finis	2	0	0	0	0	0	0	0	1	

Term	TF									DF
	Data Training					Data Testing			DF	
	D	D	D	D	D	U	U	U		
1	2	3	4	5	1	2	3			
ngajarin	1	0	0	0	0	0	0	0	1	
benci	1	0	0	0	0	0	0	0	1	
pria	1	0	0	0	0	0	0	0	1	
childfree	1	1	1	1	1	2	1	1	8	

Tabel 7 merupakan hasil perhitungan dari mencari nilai *tf* dan *df*. Proses dilakukan dengan menggunakan 8 data sampel. Data tersebut terdiri dari 5 data latih dan 3 data uji, kemudian hanya sampai pada term ke 6.

b. Perhitungan Nilai TF Normalisasi dan *idf*

Perhitungan TF normalisasi dilakukan untuk memperbaiki nilai *tf* dengan menghilangkan anomali yang disebabkan oleh perbedaan panjang dokumen. TF normalisasi dilakukan dengan frekuensi kemunculan *term* dibagi dengan panjang dokumen. Panjang dokumen, yaitu total kata yang ada pada dokumen (d). Perhitungan *idf*, dilakukan dengan persamaan (5).

c. Perhitungan TF-IDF

TF-IDF dilakukan dengan mengalikan nilai TF normalisasi dengan nilai *idf*. Hasil dari perhitungan TF-IDF, dilakukan dengan Persamaan 4. Hasil perhitungan TF-IDF dapat dilihat pada Tabel 8.

Tabel 8. Nilai TF-IDF

	TF-IDF								
	Data Training					Data Testing			DF
	D1	D2	D3	D4	D5	U1	U2	U3	
0,056443	0	0	0	0	0	0	0	0	

	D1	D2	D3	D4	D5	U1	U2	U3
0,112886	0	0	0	0	0	0	0	0
0,056443	0	0	0	0	0	0	0	0
0,056443	0	0	0	0	0	0	0	0
0,056443	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0

Tabel 8 merupakan hasil dari perhitungan mencari nilai TF-IDF. Selanjutnya hasil dari TF-IDF akan digunakan untuk melakukan klasifikasi sentimen menggunakan metode SVM.

3.4 Support Vector Machine

Proses implementasi metode Support Vector Machine akan dimulai dengan menggunakan hasil dari TF-IDF yang telah didapatkan pada proses sebelumnya. Support Vector Machine dilakukan dengan mencari *hyperplane* margin maksimum.

a. Menghitung Kernel RBF

Kernel *Gaussian* RBF merupakan kernel yang digunakan saat data tidak dapat terpisah secara linier, di mana dalam perhitungannya, RBF akan melakukan optimasi parameter *cost* dan *gamma*. Hasil dari perhitungan *kernel* RBF, dilakukan dengan persamaan (2) dapat lihat pada Tabel 9 sebagai berikut:

Tabel 9. Hasil Perhitungan Nilai Kernel RBF

	D1	D2	D3	D4	D5
D1	1	0,9704	0,9328	0,971	0,9686
D2	0,9704	1	0,9596	0,9525	0,9464
D3	0,9328	0,9596	1	0,9576	0,9538
D4	0,971	0,9947	0,9576	1	0,9929
D5	0,9686	0,9922	0,9538	0,9929	1

b. Menghitung Matrix Hessian

Hal pertama yang akan dilakukan adalah dengan menginisialisasi $\alpha_i = 0,5$. *Matrix hessian* diperoleh melalui perkalian antar kernel RBF dan kelas sentiment *y*, yang memiliki nilai 1 untuk positif, 0 untuk netral, dan -1 untuk negatif. Hasil dari perhitungan *matrix hessian* dilakukan dengan persamaan (6) dapat dilihat pada Tabel 10.

Tabel 10. Hasil Perhitungan Nilai Matrix Hessian

	D1	D2	D3	D4	D5
D1	0	0	0	0	0
D2	0	1,25	-1,2096	1,2025	-1,1964
D3	0	-1,2096	1,25	-1,2076	1,2038
D4	0	1,2447	-1,2076	1,25	-1,2429
D5	0	-1,2422	1,2038	-1,2429	1,25

c. Menghitung Sequential Training

Proses *sequential training* memiliki beberapa tahapan yang dilakukan, yakni inisialisasi variabel, yaitu α_i , γ , C , dan ϵ . Tahapan ini dilakukan dengan menginisialisasi beberapa variabel untuk diproses lebih lanjut. Proses perhitungan nilai E_i , $\delta\alpha_i$, dan α dimulai dengan tahap pertama, yaitu menghitung nilai *error rate*. Tujuan dari

perhitungan nilai *error rate* adalah untuk mendapatkan hasil *alpha* terbaru dari setiap data *training*. Hasil dari perhitungan nilai *error rate* dilakukan dengan menggunakan persamaan (7) dapat dilihat pada Tabel 11.

Tabel 11. Hasil Perhitungan Nilai *error rate*

Nilai Error	
D1	0
D2	0,0232
D3	0,0183
D4	0,0221
D5	-0,0157

Setelah nilai *error rate* pada iterasi pertama didapatkan, maka nilai *error rate* iterasi pertama pada data pertama akan digunakan untuk menghitung nilai delta alfa ($\delta\alpha_i$). Hasil dari perhitungan nilai delta alfa ($\delta\alpha_i$) dilakukan dengan menggunakan persamaan (8) dapat dilihat pada Tabel 12.

Tabel 12. Hasil Perhitungan Nilai delta alfa

$\delta\alpha_i$	
D1	0,5
D2	0,4884
D3	0,4909
D4	0,489
D5	0,5

Nilai dari delta alfa kemudian digunakan untuk mencari nilai alfa baru yang nantinya akan digunakan untuk menghitung nilai *error rate* iterasi selanjutnya. Hasil dari perhitungan alfa baru dengan menggunakan Persamaan 9.

Tabel 13. Hasil Perhitungan α_i baru

α_i	
D1	1
D2	0,9884
D3	0,9909
D4	0,989
D5	1

Tabel 13 merupakan hasil dari perhitungan mencari alfa baru untuk data *training*, yaitu D1 sampai D5. Hasil dari perhitungan nilai alfa baru selanjutnya akan digunakan untuk tahap berikutnya yaitu, menghitung nilai bias.

d. Menghitung Nilai Bias

Sebelum dilakukannya perhitungan nilai bias, tentukan nilai X^+ dan X^- terlebih dahulu dengan cara melakukan perhitungan bobot menggunakan nilai alfa baru. Hasil dari perhitungan bobot dilakukan dengan persamaan (12) dapat dilihat pada Tabel 14.

Tabel 14. Nilai Data Testing

Bobot	
D1	0,0265
D2	0,0331
D3	-0,0492
D4	0,0304
D5	0,0176

Berdasarkan tabel di atas, perhitungan nilai dari x^+ tertinggi terdapat pada data kedua dan nilai x^- terdapat pada data ketiga. Setelah nilai bobot ditentukan, langkah selanjutnya adalah mencari nilai bias dengan menggunakan Persamaan 10 yang dapat dilihat pada perhitungan sebagai berikut:

$$w \cdot x^+ = 0,3313$$

$$w \cdot x^- = -0,0492$$

$$b = \frac{1}{2}(0,331 + (-0,0492))$$

$$b = 0,0081$$

e. Menghitung Nilai Data Testing

Pada tahapan ini data uji dimasukkan ke *hyperplane* guna memperoleh klasifikasi kelas sentimen data uji. Perhitungan dilakukan dengan menghitung kernel dan nilai bobot dari data *testing* yang dimasukkan ke *hyperplane* untuk memperoleh klasifikasi kelas pada data *testing*. Untuk menghitung kernel dari data *testing*, perhitungan dilakukan sama dengan menghitung kernel dari data *training* dapat dilihat pada Tabel 15.

Tabel 15. Hasil Kernel Data Testing

	Kernel			y
U1	1	0,9842	0,9858	1
U2	0,9842	1	0,9795	-1
U3	0,9858	0,9795	1	0

Langkah selanjutnya adalah menentukan *dot product*. Setelah mendapatkan nilai *dot product*, langkah selanjutnya adalah mencari nilai dari data *testing* pada Tabel 16.

Tabel 16. Hasil Perhitungan Dot Product

Dot Product	
U1	0,0272
U2	-0,0042
U3	0,0177

Setelah mendapatkan nilai *dot product*, langkah selanjutnya adalah mencari nilai dari data *testing*. Perhitungan nilai data *testing* menggunakan persamaan (13) dapat dilihat pada Tabel 17.

Tabel 17. Hasil Nilai Data Testing

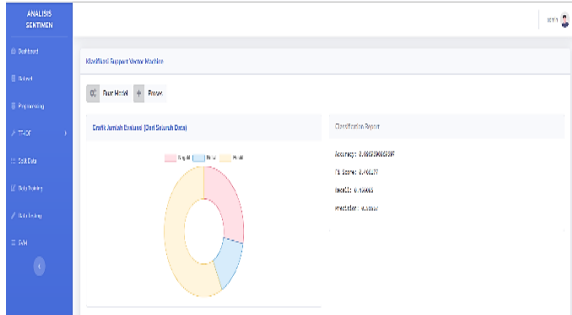
Nilai Data Testing	
U1	1,0081
U2	-0,9919
U3	1,0081

Berdasarkan hasil perhitungan dari data *testing*, maka dapat disimpulkan bahwa pada data *testing* 1 diklasifikasikan ke dalam sentimen positif, data *testing* 2 memiliki kelas sentimen negatif, dan data *testing* 3 diklasifikasikan ke dalam sentimen positif.

3.5 Confusion Matrix

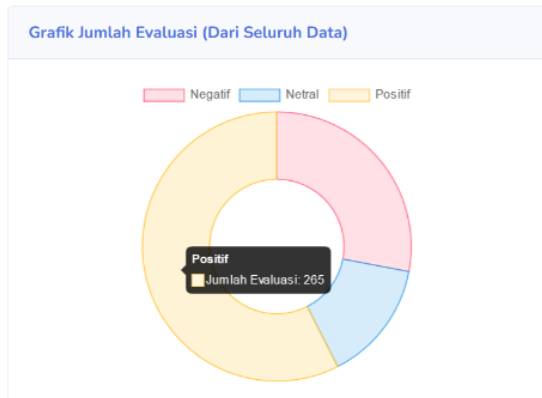
Confusion Matrix digunakan untuk mengetahui seberapa baik performa dari metode yang digunakan. Metode evaluasi yang digunakan adalah *accuracy*,

pertama diawali dengan menekan tombol “Buat Model” untuk melatih metode mengenali kelas dari label yang ada pada data *training*. Selanjutnya proses dilanjutkan dengan menekan tombol “Proses” untuk melakukan klasifikasi kelas pada data *testing*. Tampilan halaman metode Support Vector Machine dapat dilihat pada Gambar 6.



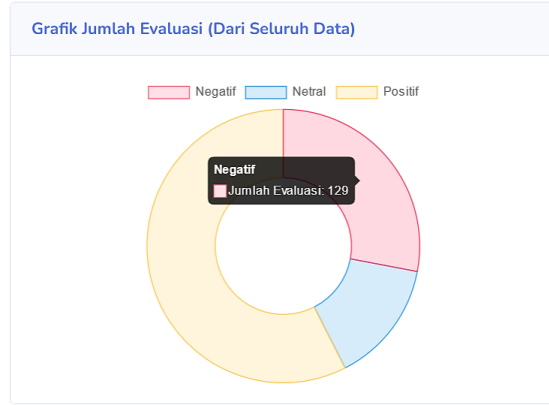
Gambar 6. Halaman Klasifikasi SVM

Berdasarkan perhitungan sistem di atas, maka dapat dilihat bahwa hasil tingkat akurasi menggunakan *confusion matrix* pada Support Vector Machine hanya mendapat nilai *accuracy* sebesar 69,69%, dengan nilai *precision* adalah 51,56%, *recall* sebesar 45,60%, dan nilai *f1-score* sebesar 46,61%. Hal ini menjelaskan bahwa penelitian mengenai analisis sentimen menggunakan metode Support Vector Machine terhadap topik *childfree* dapat dikatakan tidak memiliki performa yang baik atau cukup dalam melakukan analisis sentimen. Klasifikasi sentimen dari 461 data *tweet* yang telah diproses, komentar bersifat positif mendominasi sebesar 265 *tweets*. Hasil klasifikasi sentimen positif dapat dilihat pada Gambar 7.



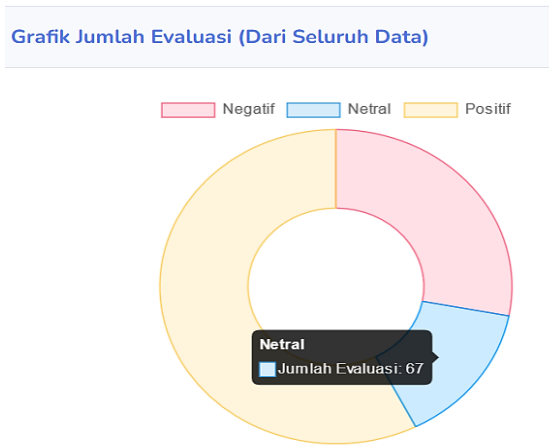
Gambar 7. Grafik Klasifikasi Sentimen Positif

Berdasarkan klasifikasi sentimen dari 461 data *tweet* yang telah diproses, komentar bersifat negatif sebesar 129 *tweets*. Hasil klasifikasi sentimen negatif dapat dilihat pada Gambar 8. Klasifikasi sentimen negatif ditandai dengan warna merah pada grafik klasifikasi sentimen.



Gambar 8. Grafik Klasifikasi Sentimen Negatif

Berdasarkan klasifikasi sentimen dari 461 data *tweet* yang telah diproses, komentar bersifat netral sebesar 67 *tweets*. Hasil klasifikasi sentimen negatif dapat dilihat pada Gambar 9. Klasifikasi sentimen negatif ditandai dengan warna merah pada grafik klasifikasi sentimen.



Gambar 9. Grafik Klasifikasi Sentimen Netral

Berdasarkan Gambar 7, 8, dan 9, dapat diambil kesimpulan perkembangan topik *childfree* sudah mulai diterima di tengah masyarakat terutama pada media sosial Twitter. Masyarakat sudah mulai berkomentar positif terhadap topik *childfree*.

4. Kesimpulan

Berdasarkan penelitian ini dapat diambil kesimpulan bahwa metode Support Vector Machine dapat diterapkan dalam melakukan analisis sentimen terhadap opini publik mengenai topik *childfree* pada media sosial Twitter. Pada penelitian ini perhitungan manual dilakukan menggunakan 8 data sampel yang terdiri dari 5 data *training* dan 3 data *testing*. Hasil dari perhitungan didapat bahwa data *testing* 1 diklasifikasikan ke dalam sentimen positif, data *testing* 2 memiliki kelas sentimen negatif, dan data *testing* 3 diklasifikasikan ke dalam sentimen positif.

Sistem analisis sentimen terhadap opini publik mengenai topik *childfree* pada media sosial Twitter

menggunakan metode Support Vector Machine dapat dibangun ke dalam sebuah sistem berbasis web menggunakan bahasa pemrograman PHP dan database MySQL. Hasil klasifikasi 461 data *tweet* yang telah diproses maka dapat disimpulkan bahwa komentar mendominasi bersifat positif sebesar 265 *tweets*, disusul dengan komentar bersifat negatif sebanyak 129 *tweets*, dan terakhir adalah komentar bersifat netral sebanyak 67 *tweets*. Hal ini menjelaskan bahwa opini masyarakat mengenai *childfree* cenderung positif dalam pembahasan topik *childfree*. Topik *childfree* sudah mulai diterima di tengah masyarakat terutama pada media sosial Twitter.

Pengujian dengan menggunakan *confusion matrix* pada metode Support Vector Machine dalam mengklasifikasi topik *childfree* dengan data sebanyak 461 data didapatkan hasil nilai *precision* sebesar 0,5156 atau 51,56%, *recall* sebesar 0,4560 atau 45,60%, *f1-score* 0,4661 atau 46,61%, kemudian untuk nilai *accuracy* sebesar 0,6969 atau 69,69%. Hal ini menjelaskan bahwa analisis sentimen menggunakan metode Support Vector Machine terhadap topik *childfree* dapat dikatakan tidak memiliki performa yang baik atau cukup dalam melakukan analisis sentimen opini public mengenai *childfree*.

Daftar Rujukan

- [1] Fitriana, F., Utami, E., & Al Fatta, H. (2021b). Analisis Sentimen Opini Terhadap Vaksin Covid - 19 pada Media Sosial Twitter Menggunakan Support Vector Machine dan Naive Bayes. *Jurnal Komtika (Komputasi dan Informatika)*, 5(1), 19–25. <https://doi.org/10.31603/komtika.v5i1.5185>
- [2] Wahyuni, W. (2022). Analisis Sentimen terhadap Opini Feminisme Menggunakan Metode Naive Bayes. *Jurnal Informatika Ekonomi Bisnis*, 148–153. <https://doi.org/10.37034/inf.v4i4.162>
- [3] Simbolon, C. A. D. (2021). Penggunaan komunikasi media sosial twitter di kalangan remaja di kecamatan cibinong, kabupaten bogor. *Jurnal Ilmu Sosial dan Ilmu Politik (JISIP)*, 10(3), 219-226.
- [4] Aisyah, I., & Hasfi, N. (2022). Opini Publik Dalam Gerakan Tagar# percumalaporpolisi Di Media Sosial Twitter. *Interaksi Online*, 10(3), 605-618 Alhaq, Z., Mustopa, A., & Santoso, J. D. (t.t.). *PENERAPAN METODE SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER*.
- [5] Siregar, D., Ladayya, F., Albaqi, N. Z., & Wardana, B. M. (2023). Penerapan Metode Support Vector Machines (SVM) dan Metode Naive Bayes Classifier (NBC) dalam Analisis Sentimen Publik terhadap Konsep Child-free di Media Sosial Twitter. *Jurnal Statistika dan Aplikasinya*, 7(1), 93-104.
- [6] Mingkase, N., & Rohmaniyah, I. (2022). Konstruksi gender dalam problematika *childfree* di sosial media Twitter. *Yinyang: Jurnal Studi Islam Gender dan Anak*, 17(2), 201-222.
- [7] Siswanto, A. W., & Neneng Nurhasanah. (2022). Analisis Fenomena *Childfree* di Indonesia. *Bandung Conference Series: Islamic Family Law*, 2(2). <https://doi.org/10.29313/bcsifl.v2i2.2684>
- [8] Fitriyah, N., Warsito, B., Asih, D., & Maruddani, I. (2020). ANALISIS SENTIMEN GOJEK PADA MEDIA SOSIAL TWITTER DENGAN KLASIFIKASI SUPPORT VECTOR MACHINE (SVM). *JURNAL GAUSSIAN*, 9(3), 376–390. <https://ejournal3.undip.ac.id/index.php/gaussian/>
- [9] Leelawat, N., Jariyapongpaiboon, S., Promjun, A., Boonyarak, S., Saengtattim, K., Laosunthara, A., Yudha, A. K., & Tang, J. (2022). Twitter data sentiment analysis of tourism in Thailand during the COVID-19 pandemic using machine learning. *Heliyon*, 8(10). <https://doi.org/10.1016/j.heliyon.2022.e10894>
- [10] Rachman, F. H. (2020). *Komputasi Bahasa Alami*. Media Nusa Creative (MNC Publishing)
- [11] Muktafin, E., Kusri, K., & Luthfi, E. T. (2020). Analisis Sentimen pada Ulasan Pembelian Produk di Marketplace Shopee Menggunakan Pendekatan Natural Language Processing. *Jurnal Eksplor Informatika*, 10(1), 32-42. <https://doi.org/10.30864/eksplor.v10i1.390>
- [12] Mukti, A., Hadiyanti, A. D., Nurlaela, A., & Panjaitan, J. (2023). Sistem Analisa Sentiment Bakal Calon Presiden 2024 Menggunakan Metode NLP Berbasis Web. *SOSCIED*, 6(1), 128-140.
- [13] Alhaq, Z., Mustopa, A., Mulyatun, S., & Santoso, J. D. (2021). Penerapan Metode Support Vector Machine Untuk Analisis Sentimen Pengguna Twitter. *Journal of Information System Management (JOISM)*, 3(1), 16-21.
- [14] Amrullah, M. S., Pane, S. F., & Fauzan, M. N. (2023). Analisis Sentimen Masyarakat Terhadap Kebijakan Polisi Tilang Manual Di Indonesia. Penerbit Buku Pedia. https://books.google.co.id/books?id=_Uq5EAAAQBAJ
- [15] Amal, M. I., Rahmasita, E. S., Suryaputra, E., & Rakhmawati, N. A. (2022). Analisis Klasifikasi Sentimen Terhadap Isu Kebocoran Data Kartu Identitas Ponsel di Twitter. *Jurnal Teknik Informatika dan Sistem Informasi*, 8(3). <https://doi.org/10.28932/jutisi.v8i3.5483>
- [16] Bourequat, W., & Mourad, H. (2021). Sentiment Analysis Approach for Analyzing iPhone Release using Support Vector Machine. *International Journal of Advances in Data and Information Systems*, 2(1), 36–44. <https://doi.org/10.25008/ijadis.v2i1.1216>
- [17] Neogi, A. S., Garg, K. A., Mishra, R. K., & Dwivedi, Y. K. (2021). Sentiment analysis and classification of Indian farmers' protest using twitter data. *International Journal of Information Management Data Insights*, 1(2). <https://doi.org/10.1016/j.jjime.2021.100019>
- [18] Maulaya, A. K. (2022). Analisis Sentimen Menggunakan Support Vector Machine Masyarakat Indonesia Di Twitter Terkait Bjorka. *Jurnal CoSciTech (Computer Science and Information Technology)*, 3(3), 495-500
- [19] Sari, E. D. N., & Irhamah, I. (2020). Analisis Sentimen Nasabah Pada Layanan Perbankan Menggunakan Metode Regresi Logistik Biner, Naive Bayes Classifier (NBC), dan Support Vector Machine (SVM). *Jurnal Sains dan Seni ITS*, 8(2), D177-D184.
- [20] Handayanto, R. T., & Herlawati. (2020). Data Mining dan Machine Learning Menggunakan Matlab dan Python. *Informatika*.
- [21] RIZKE, F. (2023). ANALISIS SENTIMEN REVIEW PENGGUNA APLIKASI PHOTOMATH DENGAN METODE SUPPORT VECTOR MACHINE (SVM).
- [22] Harieby, E., & Walid, M. (2022). TWITTER TEXT MINING MENGENAI ISU VAKSINASI COVID-19 MENGGUNAKAN METODE TERM FREQUENCY, INVERSE DOCUMENT FREQUENCY (TF-IDF). Dalam *Jurnal Mahasiswa Teknik Informatika* (Vol. 6, Nomor 2). www.kaggle.com.